CLAIMS

I/We hereby claim:

1. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising a promoter sequence operably linked to a coding sequence, said coding sequence encoding a polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM3,

said polypeptide comprising an antigenically active segment, at least five consecutive amino acids in length, of a tandem repeat sequence of the core protein of said human polymorphic epithelial mucin, which core protein is specifically bound, at the site of said segment, by monoclonal antibody SM-3, which polypeptide is specifically bound, at the site of said segment, by monoclonal antibody SM-3.

2. The vector of claim 1 wherein said segment is at least ten consecutive amino acids in length.

3. The vector of claim 1 where said tandem repeat sequence is the sequence Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly.

4. The vector of claim 1, where said polypeptide comprises a repeat sequence corresponding to a series of 20 consecutive amino acids within the 40 amino acid double repeat sequence

Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly.

5. The expression vector of claim 4 wherein said polypeptide comprises said double repeat sequence.

6. The expression vector of claim 4 wherein said polypeptide comprises three or more repeats of the repeat sequence.

7. A non-naturally occurring or isolated nucleic acid molecule, other than a whole chromosome, which comprises a coding sequence, said coding sequence encoding a polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM3,

which nucleic acid molecule specifically hybridizes under hybridizing conditions of 0.1xSSC, 0.1% SDS at 65 °C. with at least one of

I) the DNA sequence

```
5'                                        *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC-
                          *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
                                          3'
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC, or
```

II) DNA complementary to the DNA of a), i.e. of sequence

```
5'
GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT-

GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *                                       3'
```

CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT.

8. The nucleic acid molecule of claim 7 which specifically hybridizes under the same hybridization conditions with a target DNA sequence which 1) consists of three or more consecutive repeats of sequence (I), or 2) consists of three or more consecutive repeats of sequence (II).

9. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising said nucleic acid molecule of claim 7, and further comprising a promoter sequence operably linked to said coding sequence.

10. The vector of claim 9, said coding sequence being obtainable by screening a cDNA library derived from human breast cancer cell line MCF-7 for a cDNA corresponding to said coding sequence.

11. A non-naturally occurring or isolated nucleic acid molecule, other than a whole chromosome, which comprises a coding sequence, said coding sequence encoding a polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM3,

which nucleic acid molecule specifically hybridizes under hybridizing conditions of 0.1xSSC, 0.1% SDS at° C. with at least one of

I) the DNA sequence

5' *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC-

```
                          *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
                                      3'
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC, or
```

II) DNA complementary to the DNA of a), i.e. of sequence

```
5'
GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT-

GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *                                   3'
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT.
```

with each base with an asterisk immediately above it being omitted.

12. The nucleic acid molecule of claim 11 which specifically hybridizes under the same hybridization conditions with a target DNA sequence which 1) consists of three or more consecutive repeats of sequence (I), or 2) consists of three or more consecutive repeats of sequence (II), in either case with each base with an asterisk immediately above it being omitted from all repeats.

13. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising said nucleic acid molecule of claim 11, and further comprising a promoter sequence operably linked to said coding sequence.

14. The vector of claim 13, said coding sequence being obtainable by screening a cDNA library derived from human breast cancer cell line MCF-7 for a cDNA corresponding to said coding sequence.

15. A non-naturally occurring or isolated nucleic acid molecule, other than a whole chromosome, which comprises a coding sequence, said coding sequence encoding a polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM3,

and which nucleic acid molecule specifically hybridizes under hybridizing conditions of 0.1 x SSC, 0.1% SDS at 65° C. with the MCF-7-derived insert of clone pMUC10, deposited as NCIMB 40782.

16. The nucleic acid molecule of claim 15, which nucleic acid molecule specifically hybridizes under hybridizing conditions of 0.1xSSC, 0.1% SDS at° C. with at least one of

I) the DNA sequence

```
5'                                        *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC-
                          *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
                                          3'
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC, or
```

II) DNA complementary to the DNA of a), i.e. of sequence

```
5'
GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT-

GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *                                    3'
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT.
```

with each base with an asterisk immediately above it being omitted.

17. The nucleic acid molecule of claim 16, which specifically hybridizes under the same hybridization conditions with a target DNA sequence which 1) consists of three or more consecutive repeats of sequence (I), or 2) consists of three or more consecutive repeats of sequence (II), in either case with each base with an asterisk immediately above it being omitted from all repeats.

18. The nucleic acid molecule of claim 15, which nucleic acid molecule specifically hybridizes under hybridizing conditions of 0.1xSSC, 0.1% SDS at 65 °C. with at least one of

I) the DNA sequence

```
5'                                    *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC-
                        *
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
                                      3'
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC, or
```

II) DNA complementary to the DNA of a), i.e. of sequence

```
5'
GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT-

GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
  *                                    3'
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT.
```

19. The nucleic acid molecule of claim 18, which specifically hybridizes under the same hybridization conditions with a target DNA sequence which 1) consists of three or more consecutive repeats of sequence (I), or 2) consists of three or more consecutive repeats of sequence (II).

20. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising said nucleic acid molecule of claim 15, and further comprising a promoter sequence operably linked to said coding sequence.

21. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising said nucleic acid molecule of claim 16, and further comprising a promoter sequence operably linked to said coding sequence.

22. The vector of claim 21, where said polypeptide comprises a repeat sequence corresponding to a series of 20 consecutive amino acids within the 40 amino acid double repeat sequence

Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly Val Thr Ser Ala Pro Asp Thr Arg Pro Ala Pro Gly Ser Thr Ala Pro Pro Ala His Gly.

23. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising said nucleic acid molecule of claim 18, and further comprising a promoter sequence operably linked to said coding sequence.

24. An expression vector which is a recombinant DNA molecule or a purified DNA molecule, other than a whole chromosome, comprising a promoter sequence operably linked to a coding sequence, said coding sequence encoding a polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM3, said coding sequence being obtainable by screening a cDNA library derived from human breast cancer cell line MCF-7 for a cDNA corresponding to said coding sequence, said corresponding cDNA being identified by hybridization to a hybridization probe comprising

a) the DNA sequence

5'

ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC-
ACC GTG GGC TGG GGG GGC GGT GGA GCC CGG-
GGC CGG CCT GGT GTC CGG GGC CGA GGT GAC 3', or

b) DNA complementary to the DNA of a), i.e. of sequence

5'

GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT-
GTC ACC TCG GCC CCG GAC ACC AGG CCG GCC-
CCG GGC TCC ACC GCC CCC CCA GCC CAC GGT 3'.

25. A method of eliciting an immune response in a subject against an epitope specifically bound by monoclonal antibody SM-3, which comprises administering to the subject a vector according to claim 1, under conditions in which the vector directs expression of said antigen, which elicits said immune response.

26. A method of immunizing a subject against a disease characterized by the immunological presentation of an epitope specifically bound by monoclonal antibody SM-3, which comprises administering to the subject a vector according to claim 1 comprising a promoter sequence operably linked to a coding sequence, the latter encoding an antigen, under conditions in which the vector directs expression of said antigen, which elicits an immune response which is protective against such disease,
said antigen being said polypeptide comprising the core protein of a human polymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM-3.

27. The method of claim 3 in which the disease is a cancer.

28. A method of expressing an SM-3 reactive antigen in a host cell which comprises introducing into a suitable host cell a vector according to claim 1, and subjecting the cell to conditions in which the vector directs expression of said antigen, the antigen being said polypeptide comprising the core protein of a human po3lymorphic epithelial mucin, which core protein is specifically bound by monoclonal antibody SM-3.

29. The method of claim 8 in which, as a result of such expression, said antigen is accessible to the immune system of the subject.

30. The method of claim 8 in which the host cell is in a cell culture, and the expressed antigen is harvested from the cell culture.

31. A method of eliciting an immune response in a subject against an epitope specifically bound by monoclonal antibody SM-3, which comprises administering to the subject a vector according to claim 1, under conditions in which the vector directs expression of said polypeptide, which elicits said immune response.